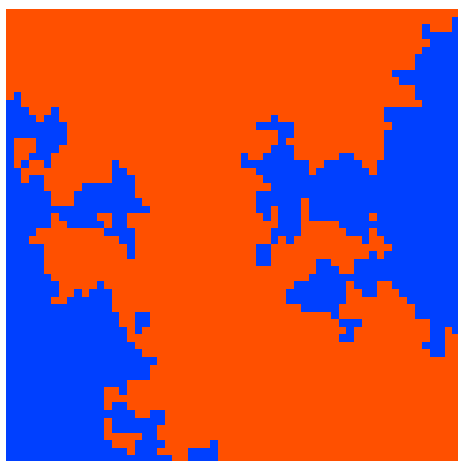


Subsurface Imaging with Support Vector Machines

Brendt Wohlberg, brendt@t7.lanl.gov
Daniel M. Tartakovsky, dmt@t7.lanl.gov
Alberto Guadagnini, alberto.guadagnini@polimi.it

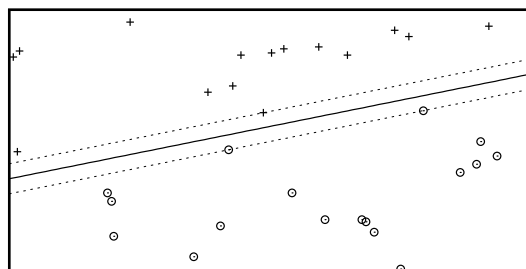
A typical subsurface environment is heterogeneous, consists of multiple materials, and is often insufficiently characterized by data. The ability to delineate geologic facies and to estimate their properties from sparse data is essential for modeling physical and biochemical processes occurring in the subsurface. Geostatistics has become an invaluable tool for estimating such properties at points in a computational domain where data are not available, as well as for quantifying the corresponding uncertainty.



Boundary between two materials in a synthetic porous medium.

One of the most popular geostatistical approaches to facies delineation employs discontinuous geostatistical models, such as Indicator Kriging [1]. In this approach, each measurement of various parameters (hydraulic conductivity and dispersivity, for example) is assigned a discrete value of an indicator function. The indicator function is treated as a random field, and Kriging is

used to estimate the values of the indicator function at points where measurements are not available. The boundary between materials is then defined as an isoline of the ensemble mean of the indicator function. The value of this isoline is determined in a way that preserves the relative volumes of each material. While successful in many applications, geostatistical methods require a number of fundamental assumptions, such as an assumption of ergodicity, whose validity is often hard to verify.

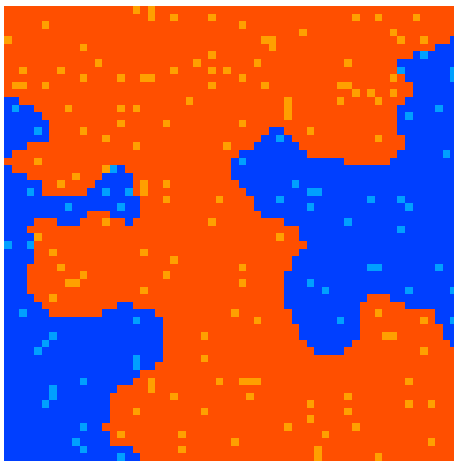


An example of a simple linear classifier applied to a problem with two classes, denoted by + and o signs. The decision boundary is represented by the solid line, and the dotted lines indicate the margin.

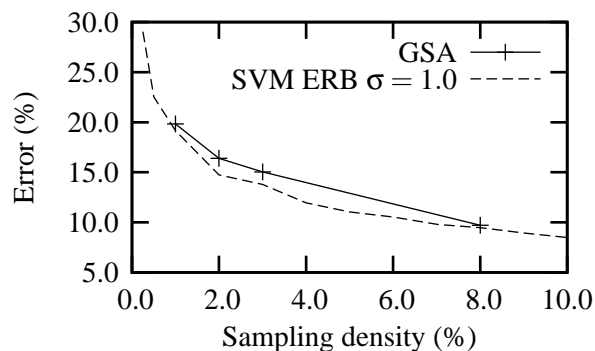
We have recently examined an alternative approach to the problem of facies delineation that uses a pattern classification technique known as the Support Vector Machine (SVM) [2, 3]. Instead of constructing an explicit probabilistic model for the underlying physical measurements or the corresponding indicator function, we treat the problem as a problem of pattern classification in the spatial domain of measurements positions. Within this framework, the boundary between materials is represented by the resulting *decision function* separating the two classes of materials. The SVM approach to this problem, motivated by Statistical Learning Theory [4], is to maximize the *margin*, indicated in the second figure by the region between the dotted lines. (While the boundary in this figure is linear, non-linear boundaries are also possible by employing *kernel* methods [3].)

Subsurface Imaging with Support Vector Machines

To demonstrate the applicability of SVMs to subsurface imaging, and to elucidate its relative advantages with respect to a geostatistical approach, we reconstruct, from a few data points selected at random, the boundary between the two heterogeneous geologic facies in the synthetic porous medium shown in the first figure. The sample points and estimated boundary are displayed in the third figure.



Boundary estimated by an SVM using the 180 indicated sample points.



Error rates for the GSA and SVM approaches.

The final figure shows a comparison of the performance of the GSA and the SVM (with the exponential radial basis kernel with radius $\sigma = 1.0$) using 20 trials for each of sampling densities. When enough measurements are available (i.e.,

when the sampling density is large enough), both methods perform equally well, with the SVM being slightly more accurate than the GSA. Two factors, however, argue strongly in favor of SVMs. First, they perform relatively well even on highly sparse data sets, on which GSA fails. Second, SVMs are highly automated, while GSAs require manual data analysis to construct spatial variograms. As a result, the GSAs are highly time consuming and depend on the subjective judgment of the practitioner.

An article on our initial experiments with linear SVMs has been published [5], and another article describing the extension to kernel SVMs is currently under review [6].

Acknowledgements

Funded by the Department of Energy under contracts W-7405-ENG-36 and the Los Alamos Laboratory Directed Research and Development (LDRD) Program.

References

- [1] A. G. JOURNAL. Nonparametric estimation of spatial distribution. *Math. Geol.*, 15(3):445 – 468, 1983.
- [2] N. CRISTIANINI AND J. SHAW-TAYLOR. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge, UK, 2000.
- [3] B. SCHÖLKOPF AND A. J. SMOLA. *Learning with Kernels*. The MIT Press, Cambridge, MA, USA, 2002.
- [4] V. N. VAPNIK. *Statistical Learning Theory*. John Wiley & Sons, Inc., New York, 1998.
- [5] D. M. TARTAKOVSKY AND B. WOHLBERG. [Delineation of geologic facies with statistical learning theory](#). *Geophys. Res. Lett.*, 31(18):L18502 (doi:10.1029/2004GL020864), 2004.
- [6] B. WOHLBERG, D. M. TARTAKOVSKY, AND A. GUADAGNINI. [Subsurface imaging with support vector machines](#). *Submitted to IEEE Trans. Geosci. Remote Sens.*, 2005.